

An Enhanced Multiway Sorting Network Based on n -Sorters

Feng Shi, Zhiyuan Yan, and Meghanad Wagh

Abstract—Merging-based sorting networks are an important family of sorting networks. Most merge sorting networks are based on 2-way or multi-way merging algorithms using 2-sorters as basic building blocks. An alternative is to use n -sorters, instead of 2-sorters, as the basic building blocks so as to greatly reduce the number of sorters as well as the latency. Based on a modified Leighton's columnsort algorithm, an n -way merging algorithm, referred to as SS-Mk, that uses n -sorters as basic building blocks was proposed. In this work, we first propose a new multiway merging algorithm with n -sorters as basic building blocks that merges n sorted lists of m values each in $1 + \lceil m/2 \rceil$ stages ($n \leq m$). Based on our merging algorithm, we also propose a sorting algorithm, which requires $O(N \log^2 N)$ basic sorters to sort N inputs. While the asymptotic complexity (in terms of the required number of sorters) of our sorting algorithm is the same as the SS-Mk, for wide ranges of N , our algorithm requires fewer sorters than the SS-Mk. Finally, we consider a binary sorting network, where the basic sorter is implemented in threshold logic and scales linearly with the number of inputs, and compare the complexity in terms of the required number of gates. For wide ranges of N , our algorithm requires fewer gates than the SS-Mk.

Index Terms—Multiway, sorting, merging

1 INTRODUCTION

Sorting is one important operation in data processing, and hence its efficiency greatly affects the overall performance of a wide variety of applications [1], [2]. Sorting networks can achieve high throughput rates by performing operations simultaneously. These parallel sorting networks have attracted attention of researchers due to increasing hardware speed and decreasing hardware cost. One of the most popular sorting algorithm is called merge-sort algorithm, which performs the sorting in two steps [2]. First, it divides the input list (a sequence of values) into multiple sublists (a smaller sequence of values) and sorts each sublist simultaneously. Then, the sorted sublists are merged as a single sorted list. The sorting process of sublists can then be decomposed recursively into the sorting and merging of even smaller sublists, which are then merged as a single sorted list. Hence, the merging operation is the key procedure for the decomposition-based sorting approach. One popular 2-way merging algorithm called odd-even merging [2] merges two sorted lists (odd and even lists) into one sorted list. In [3], a modulo merge sorting was introduced as a generalization of the odd-even merge by dividing the two sorted input lists into multiple sublists with a modulo not limited to 2. Another popular 2-way merging algorithm is bitonic merging algorithm [4]. Two sorted lists are first arranged as a bitonic list, which is then converted to obtain a sorted list. These 2-way merging algorithms employ 2-way merge procedure recursively and have a capability of sorting N

values in $O(\log^2 N)$ stages [2]. In [5], a sorting network, named AKS sorting network, with $O(\log N)$ stages was proposed. However, there is a very large constant in the depth expression, which makes it impractical. Recently, a modular design of high-throughput low-latency sorting units are proposed in [6]. However, the basic building block in these 2-way merging algorithm is a 2-sorter, which is simply a 2×2 switching element or comparator as shown in Fig. 1(a).

Instead of using 2-sorters, n -sorters can be used as basic building blocks. This was first proposed as a generalization of the Batcher's odd-even merging algorithm [7]. It was also motivated by the use of n -sorters, which sort n ($n \geq 2$) values in unit time [8], [9]. Since large sorters are used as basic building blocks, the number of sorters as well as the latency is expected to be reduced greatly. An n -way merging algorithm was first proposed by Lee and Batcher [7], where n is not restricted to 2. A version of the bitonic n -way merging algorithm was proposed by Nakatani *et al.* [10], [11]. However, the combining operation in the n -way merging algorithms still use 2-sorters as basic building blocks. Leighton proposed an algorithm for sorting r lists of c values each, represented as an $r \times c$ matrix [12]. This algorithm is a generalization of the odd-even merge-sort and named columnsort, since it merges all sorted columns to obtain a single sorted list in row order. In the original columnsort, no specific operation was provided for sorting columns and no recursive construction of sorting network was provided. In [8], a modified columnsort algorithm was proposed with sorting networks constructed from n -sorters ($n \geq 2$) [13]. However, a 2-way merge is still used for the merging process. In [14], an n -way merging algorithm, named SS-Mk, based on the modified columnsort was proposed

• Feng Shi, Zhiyuan Yan, and Meghanad Wagh are with the Department of ECE, Lehigh University, PA 18015, USA. E-mails: {fes209, yan, mduw0}@lehigh.edu.

with n -sorters as basic building blocks, where n is prime. For n sorted lists of m values each, the idea is to sort the $m \times n$ values first in each row and then in slope lines with decreasing slope rates. An improved version of the SS-Mk merge sort, called ISS-Mk, was provided in [15], where n can be any integer. We compare our sorting scheme with the SS-Mk but not the ISS-Mk, because for our interested ranges of N , the ISS-Mk requires larger latency due to a large constant.

In this work, we propose an n -way merging algorithm, which generalizes the odd-even merge by using n -sorters as basic building blocks, where $n (\geq 2)$ is prime. Based on this merging algorithm, we also propose a sorting algorithm. For $N = n^p$ input values, $p + \lceil n/2 \rceil \times \frac{p(p-1)}{2}$ stages are needed. The complexity of the sorting network is evaluated by the total number of n -sorters. The closed-form expression for the number of sorters is also derived.

Instead of 2-sorters, n -sorters ($n > 2$) are used as basic blocks in this work. This is because larger sorters have some efficient implementation. For example, for binary sorting in threshold logic, the area of an n -sorter scales linearly with the number of inputs n , while the latency stays as a constant. Hence, a smaller number of sorters and latency of the whole sorting network can be achieved. However, we cannot use arbitrary large sorters as basic blocks, since larger sorters are more complex and difficult to be implemented. Hence, the benefit of using a larger block diminishes with increasing n . We assume that the size of basic sorter $n \leq 20$ and 10 when evaluating the number of sorters and latency. Our algorithm works for any upper bound on n , and one can plug any upper bound on n into our algorithm. Asymptotically, the number of sorters required by our sorting algorithm is on the same order of $O(N \log^2 N)$ as the SS-Mk [14] for sorting N inputs. Our sorting algorithm requires fewer sorters than the SS-Mk in [14] in wide ranges of N . For instance, for $n \leq 20$, when $N \leq 1.46 \times 10^4$, our algorithm requires up to 46% fewer sorters than the SS-Mk. When $1.46 \times 10^4 < N \leq 1.3 \times 10^5$, our algorithm has fewer sorters for some segments of N 's. When $N > 1.3 \times 10^5$, our algorithm needs more sorters.

The work in this paper is different from previous works [7], [14], [15] in the following aspects:

- While the multiway merge [7] uses 2-sorters in the combining network, our proposed n -way merging algorithm uses n -sorters as basic building blocks. By using larger sorters ($n > 2$), the number of sorters as well as the latency would be reduced greatly.
- The merge-based sorting algorithms in [14], [15] are based on the modified columnsort [13], which merges sorted columns as a single sorted list in row order. Our n -way merge sorting algorithm is a direct generalization of the multiway merge sorting in [7].
- We analyze the performance of our approach by deriving the closed-form expressions of the latency and the number of sorters. We also derive the

closed-form expression of the number of sorters for the SS-Mk [14], since it was not provided in [14]. Then we present extensive comparisons between the latency and the number of sorters required by our approach and the SS-Mk [14].

- Finally, we show an implementation of a binary sorting network in threshold logic. With an implementation of a large sorter in threshold logic, we compare the performance of sorting networks in terms of the number of gates.

The rest of the paper is organized as following. In Sec. 2, we briefly review the background of sorting networks. In Sec. 3, we propose a multiway merging algorithm with n -sorters as basic blocks. In Sec. 4, we introduce a multiway sorting algorithm based on the proposed merging algorithm, and show extensive results for the comparison of our sorting algorithm and previous works. In Sec. 5, we focus on a binary sorting network, where basic sorters are implemented by threshold logic and have complexity linear with the input size, and measure the complexity in terms of number of gates. Finally Sec. 6 presents the conclusion of this work.

2 BACKGROUND

A sorting network is a feedforward network, which gives a sorted list for unsorted inputs. It is composed of two items: **switching elements (or comparators)** and **wires**. The depth of a comparator is defined to be the longest length from the inputs of the sorting network to that comparator's outputs. The **latency** of the sorting network is the maximum depth of all comparators. The network is **oblivious** in the sense that the time and location of input and output are fixed ahead of time and not dependent on the values [2]. We use the Knuth diagram in [1] for easy representation of the sorting networks, where switching elements are denoted by connections on a set of wires. The inputs enter at one side and sorted values are output at the other side, and what remains is how to arrange the switching elements. The sorting network is measured in two aspects, latency (number of stages) and complexity (number of sorters). The basic building block used by the odd-even merge [2] is a 2-by-2 comparator (compare-exchange element). It receives two inputs and outputs the minimum and maximum in an ordered way. The symbol for a 2-sorter is shown in Fig. 1(a), where x_i and y_i for $i = 1, 2$ are input and output, respectively. Similarly, an n -sorter is a device sorting n values in unit time. The symbol for an n -sorter is shown in Fig. 1(b), where x_i and y_i for $i = 1, 2, \dots, n$ are input and output, respectively, and the output satisfies $y_1 \leq y_2 \leq \dots \leq y_n$. In this work, we denote the sorted values $y_1 \leq y_2 \leq \dots \leq y_n$ by $\langle y_1, y_2, \dots, y_n \rangle$ and use n -sorters as basic blocks for sorting.

Merging-based sorting networks are an important family of sorting networks, where the merging operation is the key. There are two classes of merging algorithms, the odd-even merging [2] and the bitonic merging [4].

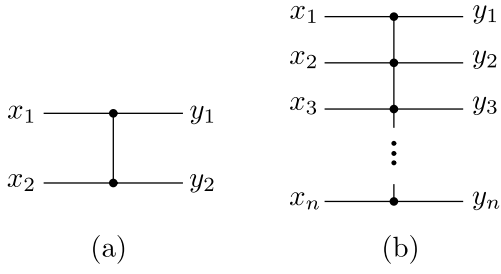


Fig. 1. (a) 2-sorter ($y_1 \leq y_2$); (b) n -sorter ($y_1 \leq y_2 \leq \dots \leq y_n$).

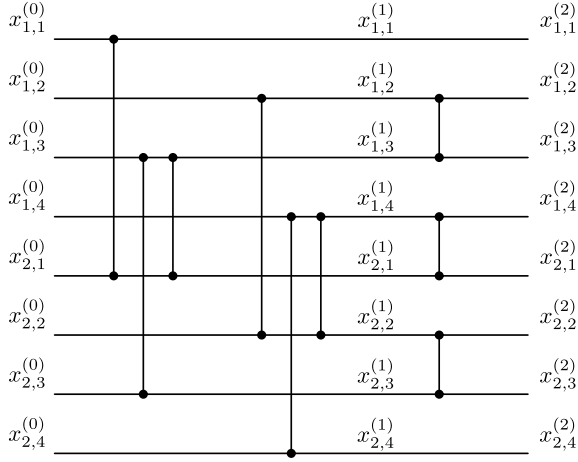


Fig. 2. The odd-even merge of two sorted lists of 4 values each using 2-sorters.

The former is an efficient sorting technique based on the divide-and-conquer approach, which decomposes the inputs into two sublists (odd and even), sorts each sublist, and then merges two sorted lists into one. Further decomposition and merging operations are applied on the sublists. An example of odd-even merging network using 2-sorters is shown in Fig. 2, where two sorted lists, $\langle x_{1,1}^{(0)}, \dots, x_{1,4}^{(0)} \rangle$ and $\langle x_{2,1}^{(0)}, \dots, x_{2,4}^{(0)} \rangle$, are merged as a single list $\langle x_{1,1}^{(2)}, \dots, x_{1,4}^{(2)}, x_{2,1}^{(2)}, \dots, x_{2,4}^{(2)} \rangle$ in two stages.

Instead of merging two lists, multiple sorted lists can be merged as a single sorted list simultaneously. An n -way merger ($n \geq 2$) of size m is a network merging n sorted lists of size m (m values) each into a single sorted list in multiple stages. This was first proposed as a generalization of the Batcher's odd-even merging algorithm. It is also motivated by the use of n -sorters, which sort n ($n \geq 2$) values in unit time [8], [9]. Since large sorters are used as basic building blocks, the number of sorters as well as the latency is expected to be reduced greatly. Many multiway merging algorithms exist in the literature [7], [8], [10]–[17]. The algorithms in [16], [17] implement multiway merge using 2-sorters. In [7], a generalization of Batcher's odd-even is introduced as shown in Fig. 3, where an n -way merger of n lists of size ud is decomposed into d n -way mergers of n sublists of size u plus a combining network. Each of the small n -way mergers is further decomposed similarly. However, the

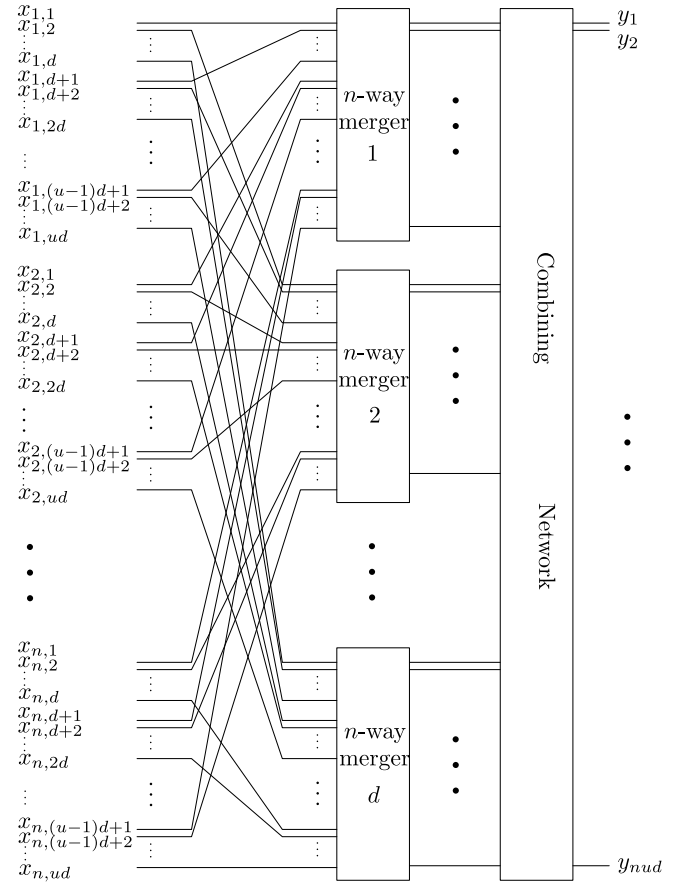


Fig. 3. Iterative construction rule for the n -way merger [7].

combining network in the merging network in Fig. 3 still uses 2-sorters as basic blocks. In [12], Leighton proposed a columnsort algorithm, which showed how to sort an $m \times n$ matrix denoting the n sorted lists of m values each. A modification of Leighton's columnsort algorithm was given in [8]. In [14], [15], merging networks with n -sorters as basic blocks are introduced based on the modified Leighton's columnsort algorithm.

In this work, we focus on multiway merge sort with binary values as inputs. Our merge sort also works for arbitrary values, which is justified by the following theorem.

Theorem 2.1 (Zero-one principle [2]). *If a network with n input lines sorts all 2^n lists of 0s and 1s into nondecreasing order, it will sort any arbitrary list of n values into nondecreasing order.*

3 MULTIWAY MERGING

In the following, we propose an n -way merging algorithm with n -sorters as basic building blocks as shown in Alg. 1. We consider a sorting network, where all iterations of Alg. 1 are simultaneously instantiated (loop unrolling). We refer to the instantiation of iteration i of Alg. 1 as stage i of the sorting network. The sorters in the last for loop in Alg. 1 consist of the last stage. Let the n sorted input lists be $\langle x_{j,1}^{(0)}, x_{j,2}^{(0)}, \dots, x_{j,m}^{(0)} \rangle$

Algorithm 1 Algorithm for n -way merging network.

Input: n sorted lists $\langle x_{j,1}^{(0)}, x_{j,2}^{(0)}, \dots, x_{j,m}^{(0)} \rangle$ for $j = 1, \dots, n$;

$i = 1$;

while $i \leq \lceil \frac{m}{2} \rceil$ **do**

for $j = 1$ to $n - 1$ **do**

 Apply $(m - i)$ -spaced sorters between lists j and $j + 1$;

end for

 Merge all $(m - i)$ -spaced sorters;

 Update n sorted lists $\langle x_{j,1}^{(i)}, x_{j,2}^{(i)}, \dots, x_{j,m}^{(i)} \rangle$ for $j = 1, \dots, n$;

$i = i + 1$;

end while

for $j = 1$ to $n - 1$ **do**

 Apply $(m - 1)$ -sorters on $m - 1$ adjacent lines with first half, $x_{j,m-k}^{(i-1)}$, from list j and second half, $x_{j+1,k}^{(i-1)}$ from list $j + 1$, where $k = 1, \dots, \frac{m-1}{2}$;

end for

Output: Sorted lists.

for $j = 1, \dots, n$. Denote the values of j -th list after stage k by $(x_{j,1}^{(k)}, x_{j,2}^{(k)}, \dots, x_{j,m}^{(k)})$. After $T = 1 + \lceil \frac{m}{2} \rceil$ stages, all input lists are sorted as a single list, $\langle x_{1,1}^{(T)}, x_{1,2}^{(T)}, \dots, x_{1,m}^{(T)} \rangle, \langle x_{2,1}^{(T)}, x_{2,2}^{(T)}, \dots, x_{2,m}^{(T)} \rangle, \dots, \langle x_{n,1}^{(T)}, x_{n,2}^{(T)}, \dots, x_{n,m}^{(T)} \rangle$.

For convenience of describing and proving our algorithm, we introduce some notations and definitions. Denote the number of zeros in the j -th list after stage i as $r_j^{(i)}$, where $i = 1, 2, \dots, \lceil \frac{m}{2} \rceil + 1$ and $j = 1, \dots, n$. A sorter is called a k -**spaced sorter** if its adjacent inputs span k other wires and each connection of the same sorter comes from different lists of m wires, where $0 \leq k \leq m - 1$. For simplicity, we arrange the sorters in the order of their first connections in each stage. Denote $\{1, 2, \dots, m\}$ as \mathbb{Z}_m . Two k -spaced sorters are said to be **adjacent** if they connect adjacent two wires, $x_{j,k}^{(i)}$ and $x_{j,k+1}^{(i)}$, respectively, for some $j \in \mathbb{Z}_m$ and $k \in \mathbb{Z}_{m-1}$. Then, our n -way merging Alg. 1 can be intuitively understood as flooding lists with zeros in descending order. The correctness of Alg. 1 can be shown by first proving the following lemmas. See the appendix for the proofs of the following lemmas and theorems.

Lemma 3.1. Apply $(m - 1)$ -spaced sorters to n lists of m values, $\langle x_{j,1}, x_{j,2}, \dots, x_{j,m} \rangle$, for $j = 1, \dots, n$. The outputs of each list are still sorted, $\langle x'_{j,1}, x'_{j,2}, \dots, x'_{j,m} \rangle$, for $j = 1, 2, \dots, n$.

For n sorted lists of m values, there are $m(m - 1)$ -spaced sorters as illustrated in Fig. 4(a). The proof of the lemma can be reduced to showing that any two wires $s, s + l \in \mathbb{Z}_m$ of each list connected by the s - and $(s + l)$ -th sorters are sorted. The simplified network is shown in Fig. 4(b). Without loss of generality, we can choose $l = 1$.

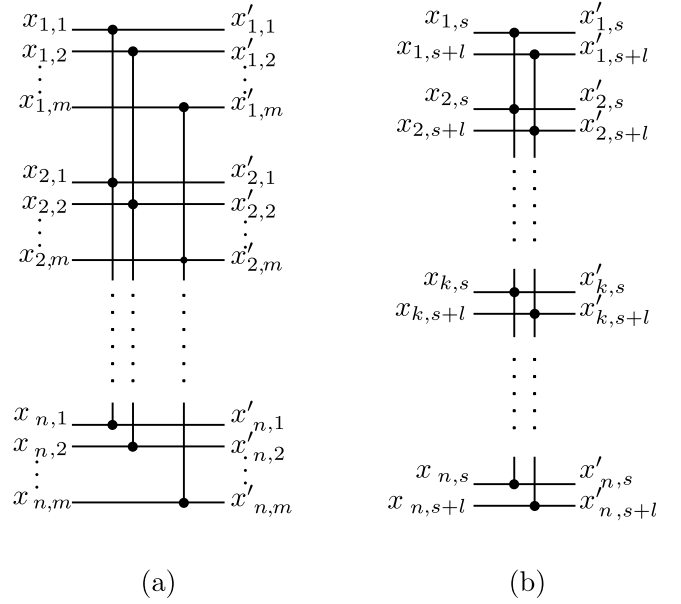


Fig. 4. The network for n sorted lists of m wires.

Lemma 3.2. In each stage of Alg. 1, there are at most four cases of adjacent two sorters as shown in Fig. 5. If m is prime, case IV is impossible.

We first show that the first connections of adjacent two sorters, $S1$ and $S2$, belong to either the same list or adjacent two lists. The same relation is true for the last connections of $S1$ and $S2$. This gives us a total of four cases as shown in Fig. 5, where $b \geq a + 1$ for Fig. 5(a)-(c), and $b \geq a$ for Fig. 5(d) such that $S1$ and $S2$ have a size of at least two.

The following theorem proves the correctness of Alg. 1.

Theorem 3.1. For a prime m in Alg. 1, all lists are self-sorted after every stage. In particular, all lists are sorted after the final stage.

The theorem can be proved by induction on i .

In Alg. 1, the latency increases linearly with $\lceil \frac{m}{2} \rceil$. When m is large, the latency is also very large. By further decomposing m into a product of small factors, we can reduce the latency significantly. In the following, we propose Alg. 2 for merging n lists of m values, where $m = n^{p-1}$ for $p \geq 2$. When m is not a power of n , we can use a larger network of $m' = n^{p'} > m$ inputs. For any q in stage i ($2 \leq i \leq p - 1$), denote the number of zeros in each new formed list after stage i as $r_{j,q}^{(i)}$, where $j = 1, \dots, n^i$. Assume two dummy lists with $r_{0,q}^{(i)} = n$ and $r_{n^i+1,q}^{(i)} = 0$ are appended to the two ends of n^i lists. The correctness of Alg. 2 can be shown by first proving the following lemma.

Lemma 3.3. In Alg. 2, the new lists in stage i with respect to q are self-sorted. The numbers of zeros of all new lists after stage i are non-increasing,

$$r_{j,q}^{(i)} \geq r_{j+1,q}^{(i)} \quad \text{for } j = 1, \dots, n^i - 1,$$

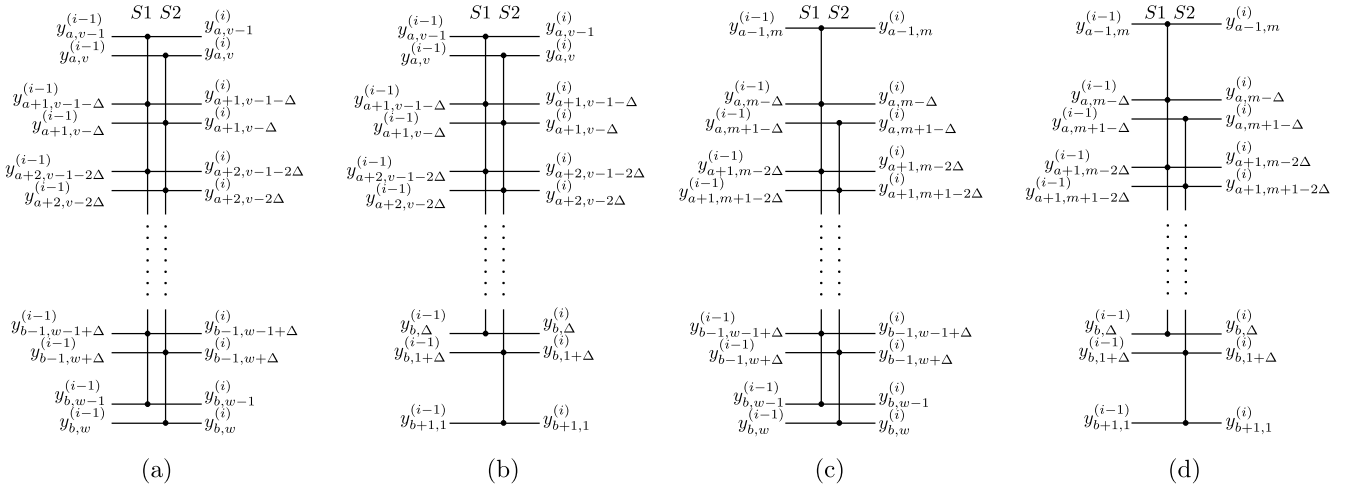


Fig. 5. Adjacent two sorters $S1$ and $S2$ in each stage of Alg. 1 can be classified into four cases. (a) Case I ($\Delta = \frac{v-w}{b-a}$); (b) Case II ($\Delta = \frac{v-1}{b-a+1}$); (c) Case III ($\Delta = \frac{m-w+1}{b-a+1}$); (d) Case IV ($\Delta = \frac{m}{b-a+2}$).

Algorithm 2 Algorithm for combining n lists of $m = n^{p-1}$ values.

Input: n sorted lists $\langle x_{j,1}^{(0)}, x_{j,2}^{(0)}, \dots, x_{j,m}^{(0)} \rangle$ for $j = 1, \dots, n$ and $m = n^{p-1}$;

$i = 1$;

for $q = 1$ to n^{p-2} **do**

 Apply Alg. 1 on $\langle x_{j,q}^{(0)}, x_{j,n^{p-2}+q}^{(0)}, x_{j,2n^{p-2}+q}^{(0)}, \dots, x_{j,(n-1)n^{p-2}+q}^{(0)} \rangle$ for $j = 1, \dots, n$ and obtain a single sorted list $\langle x_{1,q}^{(1)}, x_{1,n^{p-2}+q}^{(1)}, \dots, x_{1,(n-1)n^{p-2}+q}^{(1)}, x_{2,q}^{(1)}, x_{2,n^{p-2}+q}^{(1)}, \dots, x_{2,(n-1)n^{p-2}+q}^{(1)}, \dots, x_{n,q}^{(1)}, x_{n,n^{p-2}+q}^{(1)}, \dots, x_{n,(n-1)n^{p-2}+q}^{(1)} \rangle$;

end for

for $i = 2$ to $p-1$ **do**

for $q = 1$ to n^{p-1-i} **do**

 Group n neighboring values of $\langle x_{j,q}^{(i-1)}, x_{j,n^{p-i-1}+q}^{(i-1)}, x_{j,2n^{p-i-1}+q}^{(i-1)}, \dots, x_{j,(n-1)n^{p-i-1}+q}^{(i-1)} \rangle$ for $j = 1, \dots, n$ and denote the new lists as $\langle x_{j,q}^{(i)}, x_{j,n^{p-i-1}+q}^{(i)}, \dots, x_{j,(n-1)n^{p-i-1}+q}^{(i)} \rangle$ for $j = 1, \dots, n$;

for $k = 2$ to $\lceil \frac{n}{2} \rceil$ **do**

 Apply $(n-k)$ -spaced sorters between lists j and $j+1$;

end for

 Obtain a single sorted list $\langle x_{1,q}^{(i)}, x_{1,n^{p-i-1}+q}^{(i)}, \dots, x_{1,(n-1)n^{p-i-1}+q}^{(i)}, x_{2,q}^{(i)}, x_{2,n^{p-i-1}+q}^{(i)}, \dots, x_{2,(n-1)n^{p-i-1}+q}^{(i)}, \dots, x_{n,q}^{(i)}, x_{n,n^{p-i-1}+q}^{(i)}, \dots, x_{n,(n-1)n^{p-i-1}+q}^{(i)} \rangle$;

end for

end for

Output: Sorted list.

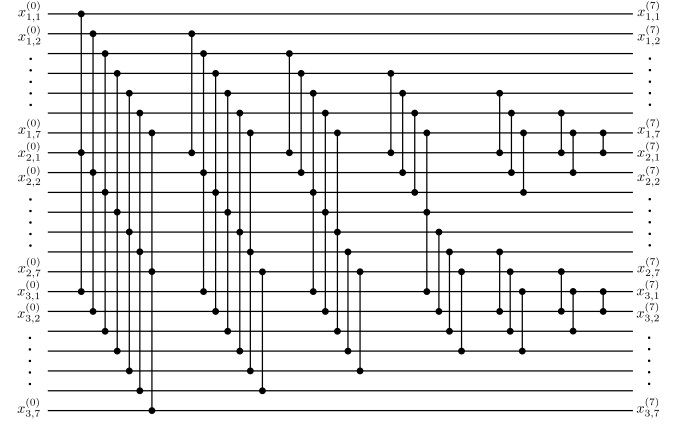


Fig. 6. A 3-way merging network of $N = 3 \times 7$ inputs implemented via 7 stages.

where $i = 2, \dots, p-1$ and $q = 1, \dots, n^{p-1-i}$. Furthermore, there are at most n consecutive lists that have between 1 and $n-1$ zeros,

$$r_{s,q}^{(i)} = n > r_{s+1,q}^{(i)} \geq \dots \geq r_{s+l,q}^{(i)} > 0 = r_{s+l+1,q}^{(i)} \quad \text{for } l \leq n,$$

where $s \geq 0$ and $s+l \leq n^i$.

See Sec. A.4 for the proof.

The following theorem proves the correctness of Alg. 2.

Theorem 3.2. Alg. 2 combines n sorted lists of $m = n^{p-1}$ values as a single sorted list.

In Alg. 2, the latency is reduced to $1 + (p-1)\lceil \frac{n}{2} \rceil$ for n sorted lists of $m = n^{p-1}$ values.

In the following, we show two examples for comparison of the two algorithms. First, a 3-way merging network of $N = 3 \times 7$ inputs via Alg. 1 is shown in Fig. 6. Then, a 3-way merging network of $N = 3 \times 9$ inputs via Alg. 2 is shown in Fig. 7. Though there are more inputs in Fig. 7 than that in Fig. 6, the latency of

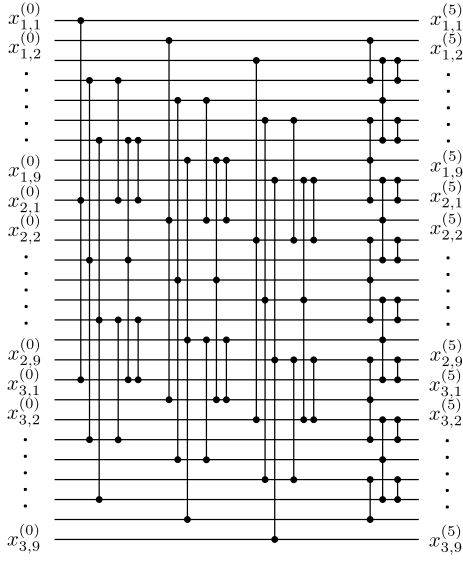


Fig. 7. A 3-way merging network of $N = 3 \times 9$ inputs implemented via 5 stages.

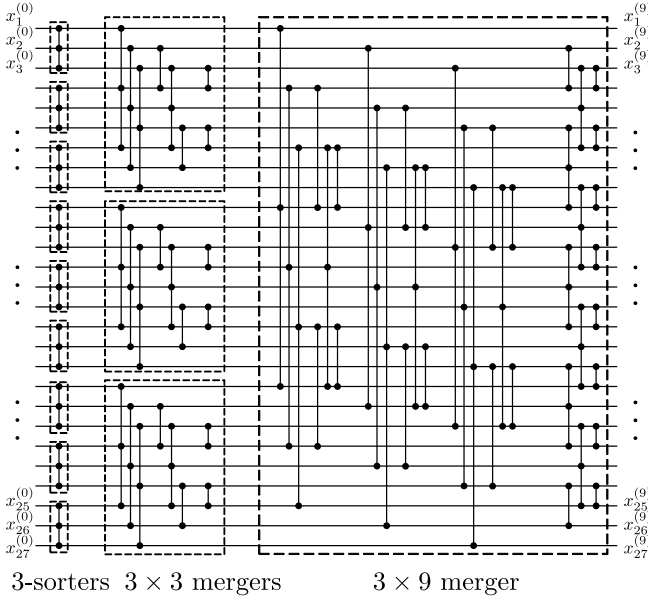


Fig. 8. A 3-way sorting network of $N = 3^3$ inputs implemented via 9 stages.

Alg. 2 is smaller due to recursive decomposition. The numbers of sorters in Figs. 6 and 7 are given by 40 and 41, respectively. For six more inputs, it requires only one more sorter in Fig. 7. Hence, Alg. 2 can be more efficient than Alg. 1 for a large m .

4 MULTIWAY SORTING

In this section, we first focus on how to construct sorting networks with n -sorters using the multiway merging algorithm in Sec. 3. Then, we analyze the latency and the number of sorters of the proposed sorting networks by deriving the closed-form expressions. We compare them with previously proposed SS-Mk in [14] but not the ISS-

Mk [15], because for our interested ranges of N , the ISS-Mk requires larger latency due to a large constant.

4.1 Multiway sorting algorithm

Based on the multiway merging algorithm in Sec. 3, we proposed a parallel sorting algorithm using a divide-and-conquer method. The idea is to first decompose large list of inputs into smaller sublists, then sort each sublist, and finally merge them into one sorted list. The sorting of each sublist is done by further decomposition. For instance, for $N = n^p$ inputs, we first divide the n^p inputs into n lists of n^{p-1} values. Then we sort each of these n lists and combine them with Alg. 2. The sorting operation of each of the n lists is done by dividing the n^{p-1} inputs into n smaller lists of n^{p-2} values. We repeat the above operations until that each of n smaller lists contains only n values, which can be sorted by a single n -sorter. The detailed procedures are shown in Alg. 3.

Algorithm 3 Algorithm for sorting $N = n^p$ values.

Input: $N = n^p$ values, $x_1^{(0)}, x_2^{(0)}, \dots, x_{n^p}^{(0)}$;
 Partition the $N = n^p$ values as n^{p-1} lists of n values each, $\langle x_{j,1}^{(0)}, x_{j,2}^{(0)}, \dots, x_{j,n}^{(0)} \rangle$ for $j = 1, \dots, n^{p-1}$;
 Apply one n -sorter on each of n^{p-1} lists and obtain $\langle x_{j,1}^{(1)}, x_{j,2}^{(1)}, \dots, x_{j,n}^{(1)} \rangle$ for $j = 1, \dots, n^{p-1}$;
for $i = 2$ to p **do**
 for $j = 1$ to n^{p-i} **do**
 Apply Alg. 1 on $\langle x_{(j-1)n+k,1}^{(i-1)}, x_{(j-1)n+k,2}^{(i-1)}, \dots, x_{(j-1)n+k,n}^{(i-1)} \rangle$ for $k = 1, \dots, n$, and obtain a single sorted list $\langle x_{j,1}^{(i)}, x_{j,2}^{(i)}, \dots, x_{j,n}^{(i)} \rangle$;
 end for
end for
Output: Sorted list.

For example, a 3-way sorting network of $N = 3^3$ inputs is shown in Fig. 8. The first stage contains 9 3-sorters. The second stage contains 3 three-way mergers with a depth of 3. The last stage contains a three-way merger with a depth of 5. The total depth is given by 9.

4.2 Latency analysis

First, we focus on the latency for sorting N values. The latency is defined as the number of basic sorters in the longest paths from the inputs to the sorted output. In Alg. 3, there are p iterations. In iteration i , there are n^i merging networks, each of which is to merge n sorted lists of n^{p-i} values. For iteration i , the latency is given by $L_{our}(n, n^{i-1}) = 1 + (i-1) \lceil \frac{n}{2} \rceil$. For a sorting network of $N = n^p$ values via Alg. 3, by summing up the latencies of all levels, we obtain the total latency

$$\begin{aligned} L_{our}(n^p) &= \sum_{i=1}^p L_{our}(n, n^{i-1}) \\ &= p + \lceil \frac{n}{2} \rceil \times \frac{p(p-1)}{2}. \end{aligned} \quad (1)$$

The closed-form expression of latency for the SS-Mk given in [14] is

$$L_{SS-Mk}(n^p) = 1 + (p-1)n + \frac{(p-1)(p-2)}{2} \lceil \log_2 n \rceil. \quad (2)$$

We compare our latency for sorting $N = n^p$ values with that for the SS-Mk in [14]. From Eqs. (1) and (2), for $N = n^p$ inputs, p should be as small as possible to obtain small latencies. In Table 1, we compare the latencies of Eqs. (1) and (2) for small p ($p = 2, 3, 4$). It is easily seen that our implementation has a smaller latency than the SS-Mk in [14] for a prime greater than 3. It is also observed that $L_{our}(2^p) = L_{SS-Mk}(2^p) = p(p+1)/2$ for $n = 2$, which is the same as the odd-even merge sort in [2].

TABLE 1

Comparison of latencies of sorting networks of $N = n^p$ inputs via the SS-Mk in [14] and our implementation.

	$p = 2$	$p = 3$	$p = 4$
[14]	$1 + n$	$1 + 2n + \lceil \log_2 n \rceil$	$1 + 3n + 3\lceil \log_2 n \rceil$
Ours	$2 + \lceil \frac{n}{2} \rceil$	$3 + 3\lceil \frac{n}{2} \rceil$	$4 + 6\lceil \frac{n}{2} \rceil$

4.3 Analysis of the number of sorters

In the following, we compare the number of sorters of our algorithms with the SS-Mk in [14]. Since the distribution of sorters for an arbitrary sorting network of N inputs is not known, we assume that any m -sorter ($m < n$) has the same delay and area as the basic n -sorter and count the number of sorters. We first derive the closed-form expression of the number of sorters for sorting N values via our Alg. 3. Since the expression of the number of sorters for the SS-Mk was not provided in [14], we also derive the corresponding closed-form expression and compare it with our algorithm. The whole sorting network is constructed recursively by merging small sorted lists into a larger sorted list. We first derive the number of sorters of a merging network of n lists of n^{p-i} values, which is given by

$$S_{our}(n, n^{p-i}) = (p-i) \cdot M_{n^{p-i}}^* + \frac{n^{p-i} - 1}{n-1} \cdot C_n^* + n^{p-i},$$

where $M_{n^{p-i}}^* = \left(1 + \frac{\lceil n/2 \rceil (\lceil n/2 \rceil - 1)}{2}\right) n^{p-i}$ and $C_n^* = (\lceil n/2 \rceil - 1)n - \frac{3\lceil n/2 \rceil (\lceil n/2 \rceil - 1)}{2} - 1$. By summing up the numbers of sorters of all mergers in all stages, we obtain the total number of sorters, which is given by

$$\begin{aligned} T_{our}(n^p) &= \sum_{i=1}^{p-1} n^{i-1} \cdot S_{our}(n, n^{p-i}) + n^{p-1} \\ &= \frac{p(p-1)}{2} \cdot M_{n^{p-1}}^* + \left[\frac{(p-1)n^{p-1}}{n-1} - \frac{n^{p-1}-1}{(n-1)^2} \right] \\ &\quad \cdot C_n^* + pn^{p-1}, \end{aligned} \quad (3)$$

As $N \rightarrow \infty$, $T_{our}(n^p)$ is on the order of $O(A_1 \frac{N \log N (\log N - \log n)}{(\log n)^2/n} + A_2 \frac{N (\log N - \log n)}{\log n} + A_3 \frac{N \log N}{n \log n})$. Similarly for the SS-Mk in [14], the number of sorters

of the merging network of n lists of n^{p-i} values each is given by

$$S_{SS-Mk}(n, n^{p-i}) = M_{n^{p-i}}^\dagger + K_{n, n^{p-i}}^\dagger + C_n^\dagger,$$

where $M_{n^{p-i}}^\dagger = \left(\frac{(n+1-\lceil n/2 \rceil)(n-\lceil n/2 \rceil)}{2} + \frac{(\lceil n/2 \rceil+1)(\lceil n/2 \rceil-2)}{2} + 2 \right) n^{p-i}$, $K_{n, n^{p-i}}^\dagger = \lceil \log_2 n^{p-1-i} \rceil n^{p-i} + (n-3)2^{\lceil \log_2 n^{p-1-i} \rceil+1}$ and $C_n^\dagger = \frac{(\lceil n/2 \rceil - 2)n - 3(\lceil n/2 \rceil+1)(\lceil n/2 \rceil-2)}{2} - \frac{(n+1-\lceil n/2 \rceil)(n-\lceil n/2 \rceil)}{2} - (n-3)$. The total number of sorters of the sorting network via the SS-Mk in [14] is given by

$$\begin{aligned} T_{SS-Mk}(n^p) &= \sum_{i=1}^{p-1} n^{i-1} \cdot S_{SS-Mk}(n, n^{p-i}) + n^{p-1} \\ &= (p-1) \cdot M_{n^{p-1}}^\dagger + \frac{n^{p-1}-1}{n-1} \cdot C_n^\dagger + n^{p-1} \\ &\quad + n^{p-1} \sum_{i=1}^{p-2} \lceil i \log_2 n \rceil \\ &\quad + \sum_{i=1}^{p-1} n^{i-1} (n-3) 2^{\lceil (p-1-i) \log_2 n \rceil+1}, \end{aligned} \quad (4)$$

As $N \rightarrow \infty$, $T_{SS-Mk}(n^p)$ is on the order of $O(B_1 \frac{N (\log N - \log n)}{(\log n)/n} + B_2 \frac{N \log N (\log N - \log n)}{n \log n} + B_3 \frac{N (\log N - \log n)}{n \log n} + B_4 \frac{N}{n})$.

According to the big-O expressions of $T_{our}(n^p)$ and $T_{SS-Mk}(n^p)$, when n is bounded, the asymptotic bounds on the number of sorters required by both our Alg. 3 and the SS-Mk in [14] are given by $O(N \log^2 N)$, which is also the asymptotical bound for the odd-even and bitonic sorting algorithms [2], [4]. When N is fixed and n increases, the first term of the big-O expressions of $T_{our}(n^p)$ and $T_{SS-Mk}(n^p)$ decreases first, then increases, and decreases to zero when $n \rightarrow N$. While other terms decrease monotonically with n . Hence, if n is not constrained, the minimum value of $T_{our}(n^p)$ and $T_{SS-Mk}(n^p)$ is one when $n = N$, meaning a single N -sorter is used.

4.4 Comparison of the number of sorters

According to the analysis of both our Alg. 3 and the SS-Mk in [14], the number of sorters for sorting $N = n^p$ inputs can be reduced by using a larger basic sorter. However, a very large basic sorter is not feasible due to some practical concerns, such as fan-in and cost. In this work, we assume that the basic sorter size is limited. For a given N , we take the total number of sorters in Eqs. (3) and (4) as a function of p with $n = N^{1/p} \leq n_b$, where n_b is the upper bound of the basic sorter size. When N is not a power of a prime, we append redundant inputs of 0's and get a larger N' such that N' is a power of a prime. Hence, we have $n' = N'^{1/p} = \lceil \lceil N^{1/p} \rceil \rceil$, where $\lceil \lceil x \rceil \rceil$ denotes the smallest prime larger than or equal to x . There exists an optimal p such that the total number of sorters is the minimum. We search for the optimal p 's for our Alg. 3 and the SS-Mk [14] using MATLAB. By plugging the optimal p 's into Eqs. (3) and (4), we obtain the total number of sorters for sorting networks of N inputs.

We compare the number of sorters for sorting networks via the Batcher's odd-even algorithm [2], our

Alg. 3, and the SS-Mk [14] for wide ranges of N . The results are shown in Fig. 9. The numbers of sorters are illustrated by staircase curves, because we use a larger sorting network for N not being a power of prime. From Fig. 9, the Batcher's odd-even algorithm using 2-sorters always requires more sorters than both our Alg. 3 and the SS-Mk in [14]. For both our Alg. 3 and the SS-Mk [14], the number of sorters is smaller for a larger n_b , meaning that using larger basic sorters reduces the number of sorters. For the comparison of the number of sorters required by our Alg. 3 and the SS-Mk [14], there are three scenarios with respect to three ranges of N . We first focus on $n_b = 10$. For $N \leq 6.25 \times 10^2$, our Alg. 3 has fewer or the same number of sorters than the SS-Mk as shown in Fig. 9. For some segments in $6.25 \times 10^2 < N \leq 3.13 \times 10^3$, our Alg. 3 has fewer sorters than the SS-Mk. For $N > 3.13 \times 10^3$, the SS-Mk in [14] needs fewer sorters. For $n_b = 20$, we have similar results. For $N \leq 1.46 \times 10^4$, our Alg. 3 has fewer or the same number of sorters than the SS-Mk as shown in Fig. 9. For some segments in $1.46 \times 10^4 < N < 1.3 \times 10^5$, our Alg. 3 has fewer sorters than the SS-Mk. For $N > 1.3 \times 10^5$, the SS-Mk in [14] needs fewer sorters.

Similarly, we compare the latency of the Batcher's odd-even algorithm, our Alg. 3, and the SS-Mk in [14]. The latencies are obtained by plugging the corresponding optimal p 's into Eqs. (1) and (2) and shown in Fig. 10 for $N \leq 2 \times 10^4$. From Fig. 10, the Batcher's odd-even algorithm using 2-sorters has the largest latency. For both our Alg. 3 and the SS-Mk [14], the latency can be reduced by having a larger n_b . The latency of our Alg. 3 is not greater than the SS-Mk for $N \leq 2 \times 10^4$ for both $n_b = 10$ and $n_b = 20$ as shown in Fig. 10. This is because our Alg. 3 tends to use large sorters, leading to less stages of sorters. We note that the latency goes up and down for some N in Fig. 10. This is because of the switching from a smaller basic sorter to a larger one to reduce the number of sorters.

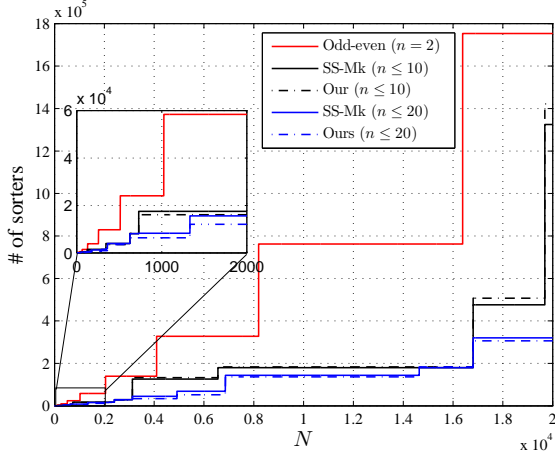


Fig. 9. Comparison of the number of sorters ($n \leq 10$ and $n \leq 20$) for sorting N inputs via the SS-Mk in [14] and our Alg. 3.

TABLE 2

Comparison of the number of sorters for sorting $N = 2^k$ inputs ($1 \leq k \leq 16$) with $n \leq 20$ via the SS-Mk in [14] and our Alg. 3.

N	SS-Mk	Ours	Rd. (%)
2	1	1	0.0
4	5	5	0.0
8	11	11	0.0
16	38	30	21.05
32	95	65	31.58
64	347	207	40.35
128	566	326	42.40
256	1250	690	44.80
512	3952	3500	11.44
1024	8287	6378	23.04
2048	15595	12039	22.80
4096	44652	33891	24.10
8192	143762	136574	5.00
16384	179631	183143	-1.96
32768	1176250	1134692	3.53
65536	1176250	1134692	3.53

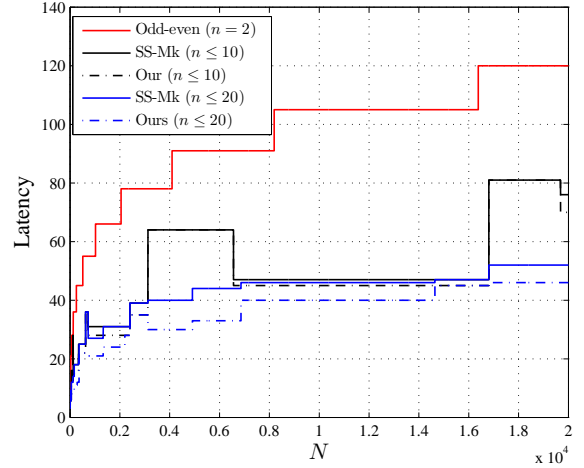


Fig. 10. Comparison of the latency for sorting N inputs with $n \leq 10$ and $n \leq 20$ via the SS-Mk in [14] and our Alg. 3.

To some researchers' interest, we also compare the number of sorters for N being a power of two. The results are shown in Table 2, where columns two and three show the numbers of sorters for the SS-Mk and our Alg. 3, respectively, and column five shows the reduction by our Alg. 3 compared with the SS-Mk [14]. For our Alg. 3, there are up to 46% fewer sorters than the SS-Mk in [14] for $N = 2^i$, for $i = 4, 5, \dots, 16$. It is also observed that a greater reduction is obtained for small p , meaning our approach is more efficient for networks with larger sorters as basic blocks.

5 APPLICATION IN THRESHOLD LOGIC

In Sec. 4.4, we assume all basic sorters in the sorting network are the same and measure the complexity by the number of sorters, since the distribution of sorters is

unknown. This would overestimate the total complexity. In this section, we focus on the threshold logic and measure the complexity by the number of threshold gates. In the following, we first briefly introduce the threshold logic, which is very powerful for computing complex functions, such as parity function, addition, multiplication, and sorting, with significantly reduced number of gates. Then, we present an implementation of a large sorter in threshold logic. Last, we compare the complexity of sorting networks in terms of the number of gates. This is a very narrow application in the sense that sorters are implemented by threshold logic and the inputs are binary values.

5.1 Threshold logic

A threshold function [18] f with n inputs ($n \geq 1$), x_1, x_2, \dots, x_n , is a Boolean function whose output is determined by

$$f(x_1, x_2, \dots, x_n) = \begin{cases} 1 & \text{if } \sum_{i=1}^n w_i x_i \geq T \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where w_i is called the *weight* of x_i and T the *threshold*. In this paper we denote this threshold function as $[x_1, x_2, \dots, x_n; w_1, w_2, \dots, w_n; T]$, and for simplicity sometimes denote it as $f = [\mathbf{x}; \mathbf{w}; T]$, where $\mathbf{x} = (x_1, x_2, \dots, x_n)$ and $\mathbf{w} = (w_1, w_2, \dots, w_n)$. The physical entity realizing a threshold function is called a threshold gate, which can be realized with CMOS or nano technology. Fig. 11 shows the symbol of a threshold gate realizing (5).

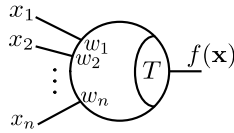


Fig. 11. Threshold gate realizing $f(\mathbf{x})$ for n inputs, x_1, x_2, \dots, x_n , with corresponding weights w_1, w_2, \dots, w_n and a threshold T .

5.2 n -sorter

Binary sorters can be easily implemented in threshold logic. In [19], a 2-by-2 comparator (2-sorter) was implemented by two threshold gates as shown in Fig. 12(a). Similarly, we introduce a threshold logic implementation of an n -sorter as shown in Fig. 12(b), where n threshold gates are required. As shown in Fig. 12, the number of gates of an n -sorter scales linearly with the number of inputs n . Hence, large sorters are preferred to be used as basic blocks. However, larger sorters are more complex and expensive to be implemented. For practical concerns, such as fan-in and cost, some limit on the size of basic sorters is assumed.

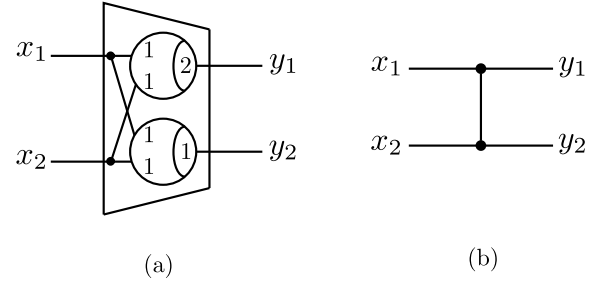


Fig. 12. Sorters implemented in threshold logic (a) 2-sorter; (b) n -sorter.

5.3 Analysis of number of gates

In the following, we assume all gates are the same and derive the total number of gates. The sorting network of N inputs is composed of multiple stages, of which each partially sorts N values. Not all values in each stage participate the comparison-and-switch operation. A simple way to count the gates is to insert buffer gates in each stage to store values without involving any sorting operation. Buffer insertion is also needed for implementation of threshold logic in some nanotechnology, where synchronization is required for correction operation. Hence, each stage contains N gates and the total number of gates is obtained by multiplying N to the latency. Note that N does not have to be a power of n . Hence, the total number of gates of our Alg. 3 and the SS-Mk [14] are simply given by

$$Q_{our}(N) = N \cdot L_{our}(N), \quad (6)$$

and

$$Q_{SS-Mk}(N) = N \cdot L_{SS-Mk}(N). \quad (7)$$

If n is bounded, the total numbers of gates in Eqs. (6) and (7) have an order of $O(N \log^2 N)$, which is the same as the order for the numbers of sorters via our Alg. 3 and the SS-Mk in [14] in Sec. 4.3.

To derive the accurate number of gates, we first derive the number of buffers added for Eqs. (6) and (7). When N is a power of prime, the number of buffers for sorting $N = n^p$ values via our Alg. 3 and the SS-Mk [14] can be easily obtained due to a regular structure. For our Alg. 3, the number of buffers is given by $G_{our}(N) = (p-1)n^{p-2} \frac{n^2+6n-5}{4} + \frac{((p-2)n^{p-1}-(p-1)n^{p-2}+1)(n+5)}{4(n-1)} + \frac{(p-1)(p-2)}{2}n^{p-1}$ for $n \neq 2$ and $G(n^p) = (p^2-p+4)2^{p-1}-2$ for $n=2$. Similarly, we derive the number of buffers for the SS-Mk in [14], which is given by $G_{SS-Mk}(N) = 2 \sum_{i=2}^p (2^{[(i-2)\log_2 n]+1} - 1)n^{p-i} + \frac{(n^{p-1}-1)(n^2-5)}{2(n-1)} + \frac{(p-1)(n-1)^2 n^{p-1}}{4}$ for $n \neq 2$ and $G(n^p) = (p^2-p+4)2^{p-1}-2$ for $n=2$. By subtracting the number of buffers from Eqs. (6) and (7), we obtain the total numbers of gates for our algorithm and the SS-Mk as shown in the following,

$$R_{our}(n^p) = n^p \cdot L_{our}(n^p) - G_{our}(n^p), \quad (8)$$

and

$$R_{SS-Mk}(n^p) = n^p \cdot L_{SS-Mk}(n^p) - G_{SS-Mk}(n^p). \quad (9)$$

Though it would overestimate the total number of gates by adding buffers. However, the asymptotic gate counts are not affected, since both $G_{our}(n^p)$ and $G_{SS-Mk}(n^p)$ have the same order of $O(N \log^2 N)$.

5.4 Comparison of the number of gates

In the following, we first compare the number of gates with consideration of buffers. Using the same idea as in Sec. 4.3, we search for the optimal p 's of Eqs. (6) and (7) using MATLAB. For $n \leq 10$ and $n \leq 20$, the numbers of gates of the SS-Mk and our two implementations are illustrated in Fig. 13. We also plot the odd-even sorting for comparison. The curves in Fig. 13 are segmented linear lines. This can be explained by Eqs. (6) and (7), which are functions of N and latency. From Fig. 13, the Batcher's odd-even algorithm using 2-sorters has more gates than both our algorithm and the SS-Mk in [14]. For both our Alg. 3 and the SS-Mk [14], the number of gates is smaller with a larger n_b , meaning that using larger basic sorters reduces the number of gates. For the comparison of the number of gates required by our Alg. 3 and the SS-Mk [14], there are also three scenarios with respect to three ranges of N . We first focus on $n_b = 10$. For $N \leq 1.68 \times 10^4$, our Alg. 3 has fewer or the same number of gates than the SS-Mk as shown in Fig. 13. For $1.68 \times 10^4 < N \leq 1.17 \times 10^5$, our Alg. 3 has the same number of gates as the SS-Mk. For $N > 1.17 \times 10^5$, the SS-Mk in [14] needs fewer gates. For $n_b = 20$, we have similar results. For $N \leq 3.71 \times 10^5$, our Alg. 3 has fewer or the same number of gates than the SS-Mk. For some segments in $3.71 \times 10^5 < N \leq 2.47 \times 10^6$, our Alg. 3 has fewer gates than the SS-Mk. For $N > 2.47 \times 10^6$, the SS-Mk in [14] needs fewer gates.

Similarly, we compare the latency of our sorting algorithm with the SS-Mk in [14]. The latencies are obtained by plugging the corresponding optimal p 's into Eqs. (6) and (7) and shown in Fig. 14 for $N \leq 2 \times 10^4$. Note that the minimization of the number of gates is essentially to minimize the latency, since each N is fixed in Eqs. (6) and (7). Fig. 14 also shows the minimal latencies of the Batcher's odd-even algorithm. All the latencies are illustrated by staircase curves. From Fig. 13, the Batcher's odd-even algorithm using 2-sorters has the largest latency. For both our Alg. 3 and the SS-Mk [14], the latency can be reduced by having a larger n_b . The latency of our Alg. 3 is not greater than the SS-Mk for $N \leq 2 \times 10^4$ for both $n_b = 10$ and $n_b = 20$ as shown in Fig. 14. This is because our Alg. 3 tends to use large basic sorters, leading to less stages.

We also compare the number of gates with buffers for N being a power of two. The numbers of gates are minimized by varying p according to Eqs. (6) and (7) for our algorithm and the SS-Mk [14]. Note the optimal p 's are different from those in Sec. 4.4. The results are shown

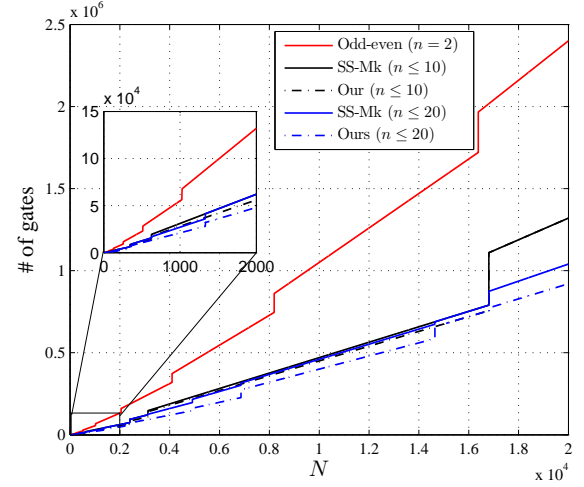


Fig. 13. Comparison of the number of gates ($n \leq 10$ and $n \leq 20$) for sorting N inputs via the SS-Mk in [14] and our Alg. 3.

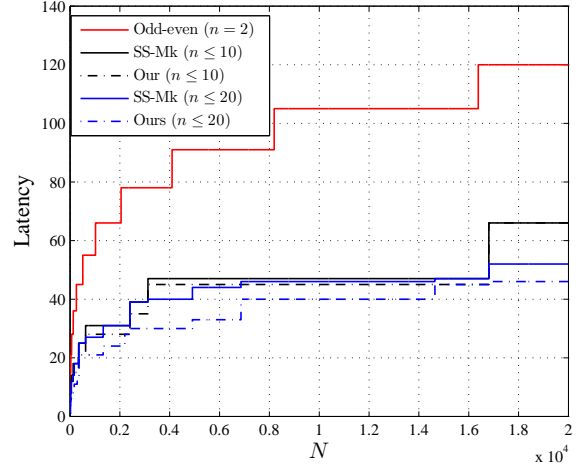


Fig. 14. Comparison of the latency ($n \leq 10$ and $n \leq 20$) for sorting N inputs via the SS-Mk in [14] and our Alg. 3.

in Table 3, where columns two to four show the numbers of gates for the SS-Mk, our Alg. 3, and the reduction of our Alg. 3, respectively, with $n \leq 20$, and columns five to seven show those with $n \leq 10$. For $n \leq 10$ and $n \leq 20$, there are up to 25% and 39% fewer gates, respectively, than the SS-Mk in [14] for $N = 2^i$ with $i = 1, 5, \dots, 16$. It is observed that fewer and the same number of gates are needed for $n \leq 20$ than for $n \leq 10$ for all $N = 2^i$ with $i = 1, 2, \dots, 16$. The reduction percentage of $n \leq 20$ is also greater than or equal to that of $n \leq 10$ for all $N = 2^i$ with $i = 1, 2, \dots, 16$ but $N = 16$. This means our sorting network takes better advantage of larger basic sorters.

For N being a power of prime, we compare the number of gates without buffers according to Eqs. (8) and (9). For $N \leq 3 \times 10^4$, we search for the same N 's for our Alg. 3 and the SS-Mk with the minimum number

TABLE 3

Comparison of the number of gates with buffers for sorting $N = 2^k$ inputs ($1 \leq k \leq 16$) with $n \leq 20$ via the SS-Mk in [14] and our Alg. 3.

N	$n \leq 20$			$n \leq 10$		
	SS-Mk	Ours	Rd. (%)	SS-Mk	Ours	Rd. (%)
2	1	1	0.00	1	1	0.00
2^2	4	4	0.00	4	4	0.00
2^3	8	8	0.00	32	32	0.00
2^4	16	16	0.00	96	80	16.67
2^5	256	192	25.00	256	192	25.00
2^6	768	512	33.33	896	768	14.29
2^7	1792	1152	35.71	2304	1920	16.67
2^8	4608	2816	38.89	4608	3840	16.67
2^9	12800	11264	12.00	12800	11264	12.00
2^{10}	27648	21504	22.22	31744	28672	9.68
2^{11}	63488	49152	22.58	63488	57344	9.68
2^{12}	163840	122880	25.00	192512	184320	4.26
2^{13}	376832	327680	13.04	385024	368640	4.26
2^{14}	770048	737280	4.26	770048	737280	4.26
2^{15}	2162688	1900544	12.12	2162688	2162688	0.00
2^{16}	4325376	3801088	12.12	4325376	4325376	0.00

of gates. The results are shown in Tables 4 and 5 for $n \leq 10$ and $n \leq 20$, respectively, where columns three and four show the numbers of gates for the SS-Mk and our Alg. 3, and column five shows the reduction of our Alg. 3. For all N 's except for $N = 7^5$, our Alg. 3 has no more gates than the SS-Mk in [14]. There are up to 13% and 23% fewer gates than the SS-Mk in [14] for $n \leq 10$ and $n \leq 20$, respectively. This means our sorting network takes better advantage of larger basic sorters. We also remark that using a larger sorter size n may reduce the number of gates for sorting $N = n^p$ inputs. For all common N 's for $n \leq 10$ in Table 4 and $n \leq 20$ in Table 5, the same number of gates is needed, since the same sorter size n is used. For all remaining N 's except for $N = 3^9$ in Table 4, there is a corresponding larger N 's in Table 5 with fewer gates. For $N = 3^9 = 19683$ in Table 4 and $N = 13^4 = 28561$ in Table 5, the latter has about 1% more gates than the former, but accounts for 45% more inputs.

6 CONCLUSION

In this work, we proposed a new merging algorithm based on n -sorters for parallel sorting networks, where n is prime. Based on the n -way merging, we also proposed a merge sorting algorithm. Our sorting algorithm is a direct generalization of odd-even merge sort with n -sorters as basic blocks. By using larger sorters ($2 \leq n \leq 20$), the number of sorters as well as the latency is reduced greatly. In comparison with other multiway sorting networks in [14], our implementation has a smaller latency and fewer sorters for wide ranges of $N \leq 1.46 \times 10^4$. We also showed an application of sorting networks implemented by linearly scaling sorters in threshold logic and have a similar conclusion that the

TABLE 4

Comparison of the number of gates without buffers for sorting $N = n^p$ inputs for $n \leq 10$ via the SS-Mk in [14] and our Alg. 3.

N	n^p	$n \leq 10$		
		SS-Mk	Ours	Rd. (%)
2	2	2	2	0.00
3	3	3	3	0.00
5	5	5	5	0.00
7	7	7	7	0.00
9	3^2	29	29	0.00
25	5^2	118	110	6.78
27	3^3	197	188	4.57
49	7^2	305	269	11.80
81	3^4	1067	998	6.47
125	5^3	1450	1315	9.31
128	2^7	2942	2942	0.00
343	7^3	5072	4728	6.78
625	5^4	13489	12140	10.00
729	3^6	22801	20411	10.48
1024	2^{10}	48126	48126	0.00
2401	7^4	63354	62254	1.74
3125	5^5	108175	97265	10.09
4096	2^{12}	278526	278526	0.00
6561	3^8	377375	330236	12.49
8192	2^{13}	655358	655358	0.00
16807	7^5	688713	704693	-2.32
19683	3^9	1443791	1259711	12.75

TABLE 5

Comparison of the number of gates without buffers for sorting $N = n^p$ inputs for $n \leq 20$ via the SS-Mk in [14] and our Alg. 3.

N	n^p	$n \leq 20$		
		SS-Mk	Ours	Rd. (%)
2	2	2	2	0.00
3	3	3	3	0.00
5	5	5	5	0.00
7	7	7	7	0.00
11	11	11	11	0.00
13	13	13	13	0.00
17	17	17	17	0.00
19	19	19	19	0.00
25	5^2	118	110	6.78
27	3^3	197	188	4.57
49	7^2	305	269	11.80
121	11^2	1117	917	17.91
125	5^3	1450	1315	9.31
169	13^2	1814	1454	19.85
289	17^2	3970	3074	22.57
361	19^2	5501	4205	23.56
625	5^4	13489	12140	10.00
729	3^6	22801	20411	10.48
1331	11^3	29107	26668	8.38
2197	13^3	54703	50763	7.20
2401	7^4	63354	62254	1.74
3125	5^5	108175	97265	10.09
4913	17^3	156812	143443	8.53
6859	19^3	239590	221052	7.74
14641	11^4	564513	562214	0.41
16807	7^5	688713	704693	-2.32
28561	13^4	1230724	1271788	-3.34

number of gates can be greatly reduced by using larger sorters.

APPENDIX A PROOFS

A.1 Proof of Lemma 3.1

Proof: The proof of the lemma can be reduced to showing that for $l > 0$ any two wires $s, s+l \in \mathbb{Z}_m$ of each list are sorted as shown in Fig. 4(b). We prove the lemma by contradiction. The inputs satisfy $x_{j,s} \leq x_{j,s+l}$ for $j \in \mathbb{Z}_n$ and $s, s+l \in \mathbb{Z}_m$. Suppose there exist $k \in \mathbb{Z}_n$ and $s, s+l \in \mathbb{Z}_m$ such that $x'_{k,s} > x'_{k,s+l}$. Since the sorter for $x_{k,s}$ $k \in \mathbb{Z}_m$ acts as a permutation of the index k , we denote such permutation of the sorter connecting wire s as $f : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$. Because f is bijection, an inverse f^{-1} exists. Then we have $x_{f^{-1}(t),s+l} \geq^a x_{f^{-1}(t),s} = x'_{t,s} \geq x'_{k,s} > x'_{k,s+l}$ for $k \leq t \leq n$, where the " \geq^a " is because the inputs are sorted and the "=" is due to the permutation. There are $n-k+1$ inputs of $x_{f^{-1}(t),s+l}$ satisfying $x_{f^{-1}(t),s+l} > x'_{k,s+l}$. However, at most $n-k$ outputs satisfy $x'_{t,s+l} > x'_{k,s+l}$ for $t \in \{k+1, k+2, \dots, n\}$, resulting in a contradiction. Hence, all lists are self sorted after applying n -sorters. \square

A.2 Proof of Lemma 3.2

Proof: First, we show that the first connections of adjacent two sorters belong to either the same list or adjacent two lists. Let (j, t_1) and $(j+l, t_2)$ be the first connections of adjacent two sorters $S1$ and $S2$, respectively, where (j, t) denotes wire t in list j . If $l > 1$, the connection of $S1$ in list $j+l-1$ should be wire m ; otherwise, $S2$ would have a valid connection in list $j+l$. For lists j to $j+l-2$, only wires m in each list are connected by $S1$, since wire m can be connected to the preceding list only by a $(m-1)$ -spaced sorter. Hence, $S1$ is the last $(m-1)$ -spaced sorter in stage 1 and $S2$ does not exist. Similarly, we can show that the last connections of adjacent two sorters $S1$ and $S2$ belong to either the same list or adjacent two lists. This gives us a total of four cases as shown in Fig. 5, where $b \geq a+1$ for Fig. 5(a)-(c), and $b \geq a$ for Fig. 5(d) such that $S1$ and $S2$ have a size of at least two.

If m is prime, no adjacent two sorters belong to case IV, which is equivalent to showing that m is a composite number if case IV in Fig. 5 exists. Assume two adjacent sorters $S1$ and $S2$ belong to case IV. Let the first connection of $S1$ be (j, m) and the last connection of $S2$ be $(j+p, 1)$. The last connection of $S1$ satisfies $(k+1)p \equiv 0 \pmod{m}$. We have $m \mid (k+1)p$. Since case IV is not possible in the first stage, we have $p < m$. Since two adjacent sorters connect two adjacent wires in at least one list, we have $p > 1$. If $k = 0$, $S1$ would connect the last and first wires of adjacent lists, respectively, in which case $S2$ does not exist. We have $1 < k+1 < m$. So m should have a proper factor dividing $k+1$ or p . Hence, m is a composite number. \square

A.3 Proof of Theorem 3.1

Proof: The theorem can be proved by induction on i . In stage 1, m -sorters are applied on corresponding wires

of all m lists. According to Lemma 3.1, the outputs of each list are sorted. Assume any two adjacent wires s and $s+1$ in list j are sorted after stage $i-1$, $x_{j,s}^{(i-1)} \leq x_{j,s+1}^{(i-1)}$ for $1 \leq j \leq n$ and $1 \leq s \leq m-1$. We will show that $x_{j,s}^{(i)} \leq x_{j,s+1}^{(i)}$ for $1 \leq j \leq n$ and $1 \leq s \leq m-1$.

According to Lemma 3.2, for a prime m , there are three cases of two adjacent sorters $S1$ and $S2$ as shown in Fig. 5(a)-(c).

- 1) For case I, let $y_{j,1}^{(i-1)}$ and $y_{j,2}^{(i-1)}$ be the two adjacent wires in list j connected by adjacent two sorters in stage $i-1$ for $a \leq j \leq b$. According to Lemma 3.1 ($n=2$), the outputs of each list are sorted.
- 2) For case II, there is an additional single wire $y_{b+1}^{(i-1)}$ connected by $S2$. If $y_{b+1,1}^{(i-1)} = 1$, we have $y_{b+1,1}^{(i)} = 1$. The last connection of $S2$ can be removed without changing the order of others in $S2$. $S1$ and the revised $S2$ reduce to case I and the outputs are sorted according to Lemma 3.1. If $y_{b+1,1}^{(i-1)} = 0$, we have $y_{b,1}^{(i-1)} = 0$. This is because they are connected by the same sorter in stage $i-1$. Then, we have $y_{a,1}^{(i)} = y_{a,2}^{(i)} = 0$, which are sorted outputs in list a . Remove $y_{b+1,1}^{(i-1)}, y_{b,1}^{(i-1)}, y_{a,1}^{(i)}$ and $y_{a,2}^{(i)}$, the remaining of $S1$ and $S2$ reduce to a smaller configuration of case II. With recursively applying the above approach, $S1$ and $S2$ either reduce to a smaller case I or a single wire, both of which gives sorted outputs.
- 3) For case III, there is an additional single wire $y_{a-1,m}^{(i-1)}$ connected by the first sorter. Similarly, the two sorters can be reduced to either a case I or a smaller configuration of case III and the outputs of two adjacent wires in each list are sorted.

Assume all lists are self-sorted after stage $i-1$, we have $x_{j,1}^{(i-1)} \leq \dots \leq x_{j,m}^{(i-1)}$ for $1 \leq j \leq n$. For stage $1 \leq i \leq \lceil \frac{m}{2} \rceil$, all wires in lists $j = 2, \dots, n-1$ have connections with some sorters. We have $x_{j,k}^{(i)} \leq x_{j,k}^{(i)}$ for $j = 2, \dots, n-1$ and $k = 1, \dots, m-1$. Hence, lists $j = 2, \dots, n-1$ are self-sorted after stage i . For list 1, $x_{1,i-1}^{(i-1)} \leq x_{2,1}^{(i-1)}$ and $x_{1,i-1}^{(i-1)} \leq x_{1,i}^{(i-1)}$, we have $x_{1,i-1}^{(i)} \leq x_{1,i}^{(i)}$. We have $\langle x_{1,1}^{(i)}, x_{1,2}^{(i)}, \dots, x_{1,i-1}^{(i)} \rangle$, since list 1 is self-sorted after stage $i-1$ and $x_{1,k}^{(i-1)} = x_{1,k}^{(i)}$ for $k = 1, \dots, i-1$. We also have $x_{1,i}^{(i)}, x_{1,i+1}^{(i)}, \dots, x_{1,m}^{(i)}$. Hence, list 1 is self-sorted after stage i , $x_{1,1}^{(i)}, x_{1,i+1}^{(i)}, \dots, x_{1,m}^{(i)}$. Due to symmetry, list n is also self-sorted after stage i , $x_{n,1}^{(i)}, x_{n,i+1}^{(i)}, \dots, x_{n,m}^{(i)}$.

To prove that the outputs of n sorted lists $\langle x_{j,1}^{(\lceil \frac{m}{2} \rceil)}, \dots, x_{j, \lceil \frac{m}{2} \rceil}^{(\lceil \frac{m}{2} \rceil)} \rangle$ for $j = 1, \dots, n$ after stage $\lceil \frac{m}{2} \rceil$ are combined as a single sorted list in stage $\lceil \frac{m}{2} \rceil + 1$, we need to show that $x_{j, \frac{m+1}{2}}^{(\lceil \frac{m}{2} \rceil + 1)} \leq x_{j, \frac{m+1}{2} + 1}^{(\lceil \frac{m}{2} \rceil + 1)}$ for $j = 1, \dots, n-1$ and $x_{j, \frac{m+1}{2} - 1}^{(\lceil \frac{m}{2} \rceil + 1)} \leq x_{j, \frac{m+1}{2}}^{(\lceil \frac{m}{2} \rceil + 1)}$ for $j = 2, \dots, n$. Since $x_{j, \frac{m+1}{2}}^{(\lceil \frac{m}{2} \rceil)} \leq x_{j, \frac{m+1}{2} + 1}^{(\lceil \frac{m}{2} \rceil)}$ and $x_{j, \frac{m+1}{2}}^{(\lceil \frac{m}{2} \rceil)} \leq x_{j+1,1}^{(\lceil \frac{m}{2} \rceil)}$, we have $x_{j, \frac{m+1}{2}}^{(\lceil \frac{m}{2} \rceil + 1)} \leq x_{j, \frac{m+1}{2} + 1}^{(\lceil \frac{m}{2} \rceil + 1)}$ for $j = 1, \dots, n-1$. Similarly, we have $x_{j, \frac{m+1}{2} - 1}^{(\lceil \frac{m}{2} \rceil + 1)} \leq x_{j, \frac{m+1}{2}}^{(\lceil \frac{m}{2} \rceil + 1)}$ for $j = 2, \dots, n$. \square

A.4 Proof of Lemma 3.3

Proof: In stage $i - 1$, there are n^{i-1} sorted lists of n values with respect to each q ($q = 1, \dots, n^{p-i}$). Since the outputs of each merging network are sorted after stage $i - 1$, we can replace each merging network by an n^i -sorter. According to Lemma 3.1, the outputs of each new formed list after stage i are sorted, $x_{j,q}^{(i)} \leq x_{j+n,q}^{(i)} \leq x_{j+(n-1)n^{p-i-1}+q}^{(i)}$ for $j = 1, \dots, n^i$. Since the corresponding wires in the new lists are connected by the same n^i -sorter in stage $i - 1$, we have $x_{j,q}^{(i)} \leq x_{j+1,q}^{(i)}$ for $j = 1, \dots, n^i - 1$. Hence, $r_{j,q}^{(i)} \geq r_{j+1,q}^{(i)}$ for $j = 1, \dots, n^i - 1$.

For $r_{s,q}^{(i)} = n > r_{s+1,q}^{(i)} \geq \dots \geq r_{s+l,q}^{(i)} > 0 = r_{s+l+1,q}^{(i)}$ for $l \leq n$, it is equivalent to prove that $x_{j+n,q}^{(i)} = 1$ if $x_{j,(n-1)n^{p-i-1}+q}^{(i)} = 1$ for $j \in \{1, \dots, n^i - n\}$. For any $q \in \{1, \dots, n^{p-1-i}\}$ in stage i , there are n^i lists of n values. Suppose $x_{j,(n-1)n^{p-i-1}+q}^{(i)} = 0$ for $j \leq s$ and $x_{s+1,(n-1)n^{p-i-1}+q}^{(i)} = 1$. If t ($t \leq s$) zeros of $x_{j,(n-1)n^{p-i-1}+q}^{(i)}$ are from the same list of the original n sorted lists, there are at most $t + 1$ zeros of $x_{j,q}^{(i)}$ from that same list. Since $x_{j,(n-1)n^{p-i-1}+q}^{(i)} = 0$ for $j \leq s$ are from at most n original lists, there are at most $s + n$ zeros in $x_{j,q}^{(i)}$, implying that $x_{s+n,q}^{(i)} = 1$. Hence, $x_{j+1,q}^{(i)} = 1$ if $x_{j,(n-1)n^{p-i-1}+q}^{(i)} = 1$ for $j \in \{1, \dots, n^i - n\}$. \square

A.5 Proof of Theorem 3.2

Proof: In stage 1, all outputs with respect to the operation of the same Alg. 1 are sorted. For any $q \in \{1, \dots, n^{p-1-i}\}$ in stage i , according to Lemma 3.3, at most n consecutive lists are not full of zeros. All preceding lists are all-zero lists and all following lists are all-one lists. Hence, the combining network in stage i is to sort n lists of n values, which is reduced to Alg. 1. In stage $p - 1$, we have $q = 1$ and the single sorted list, $\langle x_{1,q}^{(i)}, x_{1,n^{p-i-1}+q}^{(i)}, \dots, x_{1,(n-1)n^{p-i-1}+q}^{(i)}, x_{2,q}^{(i)}, x_{2,n^{p-i-1}+q}^{(i)}, \dots, x_{2,(n-1)n^{p-i-1}+q}^{(i)}, \dots, x_{n^i,q}^{(i)}, x_{n^i,n^{p-i-1}+q}^{(i)}, \dots, x_{n^i,(n-1)n^{p-i-1}+q}^{(i)} \rangle$, contains n^p values, implying all inputs are sorted as a single list. \square

REFERENCES

- [1] D. E. Knuth, "The Art of Computer Programming. Sorting and Searching, vol. III," 1973.
- [2] K. E. Batcher, "Sorting networks and their applications," in *Proc. The Spring Joint Computer Conference*. ACM, 1968, pp. 307–314.
- [3] K. J. Liszka and K. E. Batcher, "A modulo merge sorting network," in *Proc. Fourth Symposium on the Frontiers of Massively Parallel Computation*, 1992. IEEE, 1992, pp. 164–169.
- [4] K. E. Batcher, "On bitonic sorting networks," in *Proc. International Conference on Parallel Processing (ICPP)*, 1990, pp. 376–379.
- [5] M. Ajtai, J. Komlós, and E. Szemerédi, "An $o(n \log n)$ sorting network," in *Proc. The fifteenth annual ACM symposium on Theory of Computing*. ACM, 1983, pp. 1–9.
- [6] A. Farmahini-Farahani, H. J. Duwe, M. J. Schulte, and K. Comp-ton, "Modular design of high-throughput, low-latency sorting units," *IEEE Transactions on Computers*, vol. 62, no. 7, pp. 1389–1402, 2013.

- [7] D.-L. Lee and K. E. Batcher, "A multiway merge sorting network," *IEEE Transactions on Parallel and Distributed Systems*, vol. 6, no. 2, pp. 211–215, 1995.
- [8] B. Parker and I. Parberry, "Constructing sorting networks from k -sorters," *Information Processing Letters*, vol. 33, no. 3, pp. 157–162, 1989.
- [9] R. Beigel and J. Gill, "Sorting n objects with a k -sorter," *IEEE Transactions on Computers*, vol. 39, no. 5, pp. 714–716, 1990.
- [10] T. Nakatani, S.-T. Huang, B. W. Arden, and S. K. Tripathi, " k -way bitonic sort," *IEEE Transactions on Computers*, vol. 38, no. 2, pp. 283–288, 1989.
- [11] D. Lee and K. E. Batcher, "On sorting multiple bitonic sequences," in *Proc. International Conference on Parallel Processing (ICPP 1994)*, vol. 1, 1994, pp. 121–125.
- [12] T. Leighton, "Tight bounds on the complexity of parallel sorting," in *Proc. The sixteenth annual ACM symposium on Theory of Computing*. ACM, 1984, pp. 71–80.
- [13] K. J. Liszka and K. E. Batcher, "A generalized bitonic sorting network," in *Proc. International Conference on Parallel Processing (ICPP 1993)*, vol. 1, 1993, pp. 105–108.
- [14] Q. Gao and Z. Liu, "Sloping-and-shaking," *Science in China Series E: Technological Sciences*, vol. 40, no. 3, pp. 225–234, 1997.
- [15] L. Zhao, Z. Liu, and Q. Gao, "An efficient multiway merging algorithm," *Science in China Series E: Technological Sciences*, vol. 41, no. 5, pp. 543–551, 1998.
- [16] R. Drysdale, III and F. H. Young, "Improved divide sort merge sorting networks," *SIAM Journal on Computing*, vol. 4, no. 3, pp. 264–270, 1975.
- [17] D. C. Van Voorhis, "An economical construction for sorting networks," in *Proceedings of the May 6-10, 1974, national computer conference and exposition*. ACM, 1974, pp. 921–927.
- [18] S. Muroga, *Threshold Logic and Its Applications*. New York: WILEY-INTERSCIENCE, 1971.
- [19] V. Beiu, J. Peperstraete, and R. Lauwereins, "Enhanced threshold gate fan-in reduction algorithms," in *Proceedings of the third international conference on Young computer scientists*. Tsinghua University Press, 1993, pp. 339–342.